WHAT DO REGRESSION ANALYSES OF INTER-STRESS INTERVAL DURATION REALLY MEASURE?

Andreas Windmann[1], Juraj Šimko[2], Petra Wagner[1]

[1]Bielefeld University, Germany [2]University of Helsinki, Finland
[1]andreas.windmann@uni-bielefeld.de, petra.wagner@uni-bielefeld.de, [2]juraj.simko@helsinki.fi

ABSTRACT

Language-specific differences in inter-stress interval duration expressed as a function of the number of component syllables have been taken as evidence for oscillatory mechanisms in speech timing. Using corpus data and computer simulations, we provide support for the alternative hypothesis that cross-linguistic differences in the distribution of stressed syllables relative to word boundaries may account for the observed effect.

Keywords: Speech Timing, Speech Rhythm

1. INTRODUCTION

This paper proposes a possible explanation for empirical results on the duration of the interval between the onsets of two consecutive stressed syllables (inter-stress interval; ISI) in running speech expressed as a function of the number of syllables in this interval. In a regression analysis of this kind, [8] found an interesting cross-linguistic pattern: while ISI duration was found to increase linearly with syllable count in all investigated languages at a rate of approximately 100 ms per added syllable (contrary to claims about isochrony of ISI in "stress-timed" languages, cf. [13, 1]), intercept estimates differed in an interesting way, clustering narrowly around 200 ms in "stress-timed" languages and around 100 ms in "syllable-timed" languages.

An elegant explanation of this fact was proposed by [11]. In their mathematical model, the empirical difference is explained by coupling between periodic oscillators at the syllabic and the ISI level, such that they entrain to a harmonic frequency pattern. This model predicts ISI duration to depend linearly on the number of component syllables, with an added positive constant. Differences in this constant term arise from varying the nature of the coupling between the two oscillators: a larger intercept term will be generated if the stress-group oscillator is made more dominant, i.e., more "reluctant" to depart from its periodic natural frequency, while the syllabic oscillator has to adjust its period more strongly. In particular, the value of the responsible model parameter can be derived from the ratio between regression intercept and slope estimated on the empirical data. [8]'s results suggest that this *coupling strength ratio*, denoted *r*, be set to 2 for "stress-timed" and to 1 for "syllable-timed" languages. Thus, the coupled oscillator model is compatible with a weaker, continuous version of the isochrony hypothesis: a stronger *tendency* towards isochrony of ISI in "stress-timed" languages generates the empirically observed difference.

Despite the impressive formal elegance of this model, alternative explanations remain available. [8] himself suggested that the non-zero intercepts in his analysis result primarily from the greater duration of stressed compared to unstressed syllables. The language-specific differences, then, may simply be a consequence of greater stressed/unstressed syllable duration ratios in "stress-timed" compared to "syllable-timed" languages. This assumption converges with differences in language structure: "stress-timed" languages are characterized by features such as complex syllables attracting stress and reduction phenomena in unstressed syllables to a greater degree than "syllable-timed" languages [5].

However, differences in stressed/unstressed ratios cannot be the only explanation for [8]'s result: [12] found an intercept estimate of 104 ms despite an average stressed-unstressed duration difference of only 13 ms in a regression analysis of ISI duration in Finnish. They argue that this points towards "a real dependence of syllable rate on stress group size", as predicted by the coupled oscillator model. In this paper, we will investigate the alternative hypothesis that cross-linguistic differences in the alignment of stressed syllables and word boundaries may be responsible for regression results on ISI duration. To this end, we will report experiments on artificial speech corpora generated with a statistical "toy model" that incorporates minimal assumptions about temporal and distributional properties of the investigated languages. While this model is very crude, it allows us to isolate the effect of the alignment of stressed syllables and word boundaries

on speech timing. The goal of the study is to provide a proof-of-concept demonstration, showing that all else being equal, differences in the alignment of stressed syllables and word boundaries predict tendencies in the observed direction. The paper will be closed by a discussion of our findings in the light of current theoretical debates about speech timing.

2. ANALYSES

A recent investigation [21] has shown that the finding of polysyllabic shortening at the ISI level (which is also predicted by coupled oscillator models) as observed in corpus analyses may be an artifact of the alignment of stressed syllables and word boundaries. The relevant result is shown in the left panel of Figure 1, which graphs the percentage of stressed syllables that are word-final as a function of syllable count in the ISI computed on a large corpus of English [3]. It is apparent that stressed syllables in longer ISI are less likely to appear in word-final position, and if one assumes that word-final syllables are lengthened, this distributional fact would account for apparent polysyllabic shortening effects, at least on stressed syllable duration.

One may hypothesize that this pattern should also have an effect on ISI duration itself. In particular, it may explain cross-linguistic differences, as it may vary with stress assignment rules in different languages. In this study, we will consider two extreme cases along with English: in French, with its supposedly post-lexical assignment of prominence to the final syllable in an accentual phrase, all "stressed" syllables are necessarily word-final, regardless of syllable count in the ISI, as shown in the middle panel of Figure 1 [18]. There may be exceptions to this pattern in the case of phrase-initial accents [2], but the occurrence of this phenomenon should be randomly distributed with regard to ISI length. French is also a prototypical "syllable-timed" language.

The other extreme case is represented by Finnish, where stress always falls on the initial syllable of a word [16], such that the only possibility for a stressed syllable to occur word-finally is in a monosyllabic word. The resulting distribution of word-final stressed syllables would presumably resemble that in the right panel of Figure 1: Monosyllabic ISI in Finnish occur only if a monosyllabic word is followed by any other word (which, in theory, necessarily has initial stress), hence stressed syllables in monosyllabic ISI are always word-final. In longer ISI, i.e., sequences of a stressed and one or more unstressed syllables, the stressed syllable cannot be word-final, as any unstressed syllable that follows a stressed syllable must also fall within the same word. This is most likely an overstatement, as there may be cases of de-stressing in running speech, but the resulting distribution should at least be markedly more extreme than in English, where polysyllabic words with final stress do occur.

Figure 1: (Percentage of stressed syllables that are word-final by syllable count in the ISI in English (computed on MARSEC corpus), French and Finnish (hypothetical distributions).

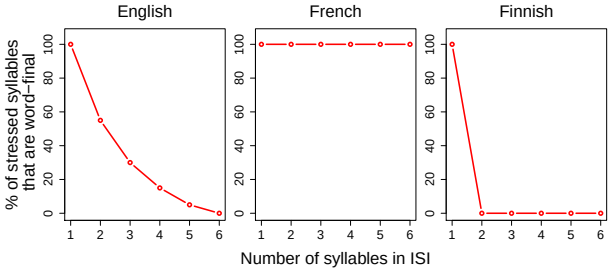


Table 1 shows results of regression analyses of ISI duration on syllable count in the three languages. The data come from the Aix-MARSEC corpus [3] for English, and from the C-PROM corpus [4] for French. Following [9] and [18], the ISI in French, in contrast to English was defined as the interval between the *offset* of a "stressed" syllable and the offset of the following one. Linear regression models were fitted to durations of utterance-medial ISI by syllable count. The Finnish result was taken from [12]. Intercept and coupling strength ratio r for English, while lower than commonly reported in the literature, are markedly higher than in the other languages, as predicted by an oscillatory account that assumes a greater relative dominance of the stress oscillator in English than in French and Finnish.

Table 1: *Regression models of ISI duration by syllable count in English, French and Finnish.*

Language	Int. (ms)	Slope (ms)	r	R^2
English	143	123	1.16	0.49
French	122	142	0.85	0.78
Finnish [12]	104	145	0.71	0.73

In order to investigate whether the patterns shown in Figure 1 provide an alternative explanation for the results in Table 1, we conducted a study on simulated speech data generated using a minimal statistical "toy model" of speech timing: syllable duration data from the three languages were simulated by randomly drawing numbers from log-normal distributions. The log-normal distribution has been

shown to provide a good approximation of the typically positively skewed distributions of speech segment durations [15]. Three categories of syllables were modeled: word-final stressed syllables were drawn from a distribution with a mean $\mu = 265$ ms and a standard deviation $\sigma = 105$ ms, non-word-final stressed syllables from a distribution with $\mu = 178$ ms and $\sigma = 75$ ms, and unstressed syllables from a distribution with $\mu = 147$ ms and $\sigma = 40$ ms. These values were derived from the MARSEC corpus. They were deliberately used for all languages, so as to isolate the effect of the distribution of word-final stressed syllables on regression results. The simulated syllable durations were combined into ISI as shown in Table 2, so as to match the proportions of word-final and non-final stressed syllables shown in Figure 1. For example, the "English" corpus comprised 200 tri-syllabic ISI, 70 of which were generated by adding a syllable duration from the word-final stressed distribution to two unstressed durations, while the other 130 tri-syllabic ISI durations were generated by adding two unstressed durations to a stressed non-final duration. The assumption of equally many observations for ISI of all lengths is of course also an oversimplification.

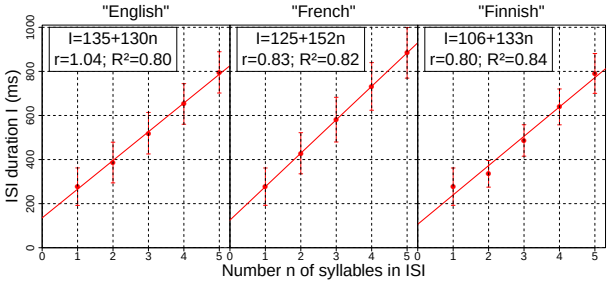
Table 2: *Counts of stressed non-final (NF) and stressed final (F) syllables by ISI length in the simulated corpora (the number of unstressed syllables for an ISI of length n is* $200 \times (n-1)$*).*

		Number n of syllables in ISI				
Lang.	Pos.	1	2	3	4	5
"En"	F	200	110	70	30	10
	NF	0	90	130	170	190
"Fi"	F	200	0	0	0	0
	NF	0	200	200	200	200
"Fr"	F	200	200	200	200	200
	NF	0	0	0	0	0

Linear regression models were fitted to ISI duration by syllable count on the simulated data, and the "coupling strength ratio" r, i.e., the ratio between intercept and slope, was computed. This procedure was repeated 500 times, so as to produce stable estimates. Figure 2 shows mean durations and regression coefficients computed over 500 simulations.

The simulation result confirms our hypothesis for the comparison between English and the other two languages: all else being equal, the difference in distribution of word-final stressed syllables alone predicts a markedly higher coupling strength ratio for English. For the comparison with French, the explanation lies in the statistical tendency of stressed syllables to be progressively shorter in longer ISI (even though ISI length is not the relevant variable as far as timing processes are concerned), which counteracts the increase in ISI duration caused by the addition of unstressed syllables and decreases the regression slope. In Finnish, the abrupt difference between mono- and bisyllabic ISI (100% vs. 0% word-final stressed syllables) makes for a nonlinearity that results in a lower intercept estimate, and, hence, coupling strength ratio compared to English. Interestingly, graphical presentation of Finnish ISI duration means in [12] reveals a very similar nonlinearity.

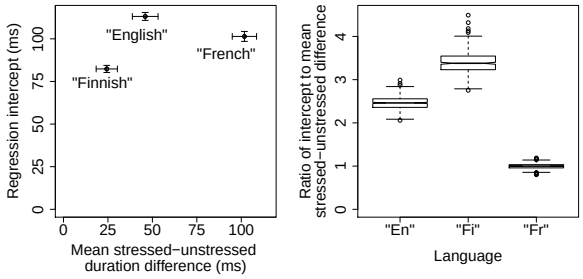
Figure 2: Regressions on simulated English, French and Finnish data (means, standard deviations and regression coefficients averaged over 500 simulation runs).



Could [12]'s result of a substantial positive regression intercept despite minimal stressed-unstressed syllable duration differences also be a consequence of the particular language structure of Finnish? In order to answer this question, we re-ran the simulation with new durational parameters, which were chosen so as produce a small average stressed-unstressed duration difference, as in [12]'s data: μ(stressed word-final)=265 ms, μ(stressed non-word-final)=175 ms, μ(unstressed)=170 ms. Standard deviations were the same as above. Again, the same duration parameters were used for all three languages. Results are shown in Figure 3. The left panel graphs simulated regression intercepts as a function of the difference between mean stressed and unstressed syllable duration over 500 simulation runs for each language, and the right panel shows the ratio of both quantities by language.

The left panel of Figure 3 shows that the Finnish-like language structure indeed makes for a substantial positive regression intercept despite a small stressed-unstressed duration difference. This becomes even more evident in the right panel of the Figure: in the Finnish simulation, the regression intercept is on average more than three times as large as the difference between mean stressed and unstressed syllable duration, a substantially larger ratio

Figure 3: Left panel regression intercept by mean stressed-unstressed duration difference in simulated English, French and Finnish data. Right panel: ratio between both variables by language.



than in the other languages. This ratio is still not sufficient to fully explain the outcome of [12], but this may be due to other structural factors not considered in the simulation, such as the distribution of pitch accents, secondary stress, or final lengthening in *unstressed* syllables. The crucial result is that the particular distributional properties of Finnish impose a tendency in the observed direction.

The ratio between intercept and stressed-unstressed difference is markedly lower in the other languages – in particular, for French, where the proportion of word-final stressed syllables is independent of ISI length, it is exactly one. English, in this analysis, occupies a middle position between French and Finnish, which is not surprising, given that its distributional pattern of word-final stressed syllables is also halfway between the extremes French and Finnish (cf. Figure 1). The simulation thus makes a prediction that we can verify on our corpora – the regression intercept should be markedly higher than the difference between mean stressed and unstressed syllable duration in English, while both quantities should be about the same in French. This prediction is partially confirmed by our data: analysis of the MARSEC corpus yields a regression intercept of 143 ms for English and a difference between mean stressed and unstressed syllable duration of 82 ms. In the French C-PROM data, the regression intercept is 122 ms and the stressed-unstressed difference is 94 ms. While the ratio between both numbers is thus not equal to 1 in French, it is at least substantially lower than in English (French: 1.30, English: 1.74).

3. DISCUSSION

Our results show that differences in the relative frequency and distribution of stressed syllables that are subject to word-final lengthening in different languages may contribute to regression results by [8]. Such differences thus provide an alternative to coupled oscillators in accounting for these results. Our very coarse simulations did obviously not provide perfect matches to empirical results, and there may be many other structural factors not implemented in our toy model that influence the relationship between ISI duration and the number of component syllables. However, our analysis has created a *ceteris paribus* condition, showing that all else being equal, the structural differences we modeled trigger tendencies in the observed direction.

Do our findings also account for the lower coupling strength ratios for languages such as Italian, Spanish and Greek compared to English reported by [8]? As for Italian, [10] reports that only about 2% of all lexemes in Italian have final stress, which would suggest that the distribution of word-final stressed syllables in Italian may follow a Finnish-like pattern. In Spanish, word-final stressed syllables are not vanishingly rare [7], but stressed-unstressed duration ratios are markedly lower than in English [14]. Of course, word-final lengthening of stressed syllables may also be weak or absent in these languages – for example, [6] report that at least in open syllables, stressed vowels are *shorter* word-finally than elsewhere in Italian. As for the languages in our study, evidence for word-final lengthening has been supplied by [17] as well as by our own analysis for English and by [16] for Finnish. For French, the question is moot, as all "stressed" syllables in French are word-final.

Our results are compatible with the domain-and-locus approach to speech timing [19, 20], which argues that suprasegmental speech timing is confined to localized lengthening effects at the heads and edges of prosodic domains, and is not based on underlying periodicities as seems to be implied by oscillatory models. In accordance with this approach, alternative explanations for apparent polysyllabic shortening effects in various prosodic constituents have been proposed [19, 17, 21]. The present study complements these findings by suggesting that regression results on ISI duration [8, 12] can be accounted for in a strictly localized model of speech timing, and do not necessarily require explanations based on underlying periodicities.

4. ACKNOWLEDGEMENTS

We would like to thank the authors of [3] and [4] for making their corpora available to the research community. The first author is funded by the Bielefeld graduate school of linguistics and literary studies.

5. REFERENCES

[1] Abercrombie, D. 1967. *Elements of general Phonetics*. Edinburgh: Edinburgh University Press.

[2] Astésano, C., Bard, E. G., Turk, A. 2007. Structural influences on initial accent placement in french. *Language and Speech* 50(3), 423–446.

[3] Auran, C., Bouzon, C., Hirst, D. 2004. The Aix-MARSEC project: an evolutive database of spoken British English. *Proceedings of Speech Prosody 2004* Nara, Japan.

[4] Avanzi, M., Simon, A.-C., Goldman, J.-P., Auchlin, A. 2010. C-PROM: An annotated corpus for French prominence study. *Proceedings of Speech Prosody 2010* Chicago.

[5] Dauer, R. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11(1), 51–62.

[6] d'Imperio, M., Rosenthall, S. 1999. Phonetics and phonology of main stress in Italian. *Phonology* 16(01), 1–28.

[7] Eddington, D. 2000. Spanish stress assignment within the analogical modeling of language. *Language* 92–109.

[8] Eriksson, A. 1991. *Aspects of Swedish speech rhythm*. PhD thesis University of Gothenburg.

[9] Fant, G., Kruckenberg, A., Nord, L. 1991. Durational correlates of stress in Swedish, French, and English. *Journal of Phonetics* 19(1), 351–365.

[10] Krämer, M. 2009. Main stress in Italian nonce nouns. In: Torck, D., Wetzels, L., (eds), *Romance Languages and Linguistic Theory 2006: Selected Papers from 'Going Romance'* volume 303. John Benjamins Publishing 127–142.

[11] O'Dell, M., Nieminen, T. 1999. Coupled oscillator model of speech rhythm. *Proceedings of ICPhS 1999* San Francisco. 1075–1078.

[12] O'Dell, M., Nieminen, T. 2001. Speech rhythms as cyclical activity. *Proceedings of the XXIst Finnish Phonetics Symposium* Turku. 159–168.

[13] Pike, K. 1945. *The Intonation of American English.* Ann Arbor: University of Michigan Press.

[14] Prieto, P., Vanrell, M. d. M., Astruc, L., Payne, E., Post, B. 2012. Phonotactic and phrasal properties of speech rhythm. Evidence from Catalan, English, and Spanish. *Speech Communication* 54(6), 681–702.

[15] Rosen, K. M. 2005. Analysis of speech segment duration with the lognormal distribution: A basis for unification and comparison. *Journal of Phonetics* 33(4), 411–426.

[16] Suomi, K. 2007. On the tonal and temporal domains of accent in Finnish. *Journal of Phonetics* 35(1), 40–55.

[17] Van Santen, J. P. 1992. Contextual effects on vowel duration. *Speech Communication* 11(6), 513–546.

[18] Wenk, B. J., Wioland, F. 1982. Is French really syllable-timed? *Journal of Phonetics* 10, 193–216.

[19] White, L. 2002. *English speech timing: a domain and locus approach*. PhD thesis University of Edinburgh.

[20] White, L. 2014. Communicative function and prosodic form in speech timing. *Speech Communication* 63, 38–54.

[21] Windmann, A., Šimko, J., Wagner, P. 2014. Probing theories of speech timing using optimization modeling. *Proceedings of Speech Prosody 2014* Dublin, Ireland. 346–350.